LEVEL II

(12)

DTIC
SELECTED
AUG 0 1 1980

E

DEPARTMENT OF STATISTICS

The Ohio State University

OSU

COLUMBUS, OHIO

80 7 29 027
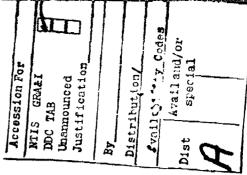
LEVEL II

(12)

HETEROSCEDASTICITY*

by

Edward J. Dudewicz

DTIC
ELECTE
AUG 0 1980
E

Technical Report No. 209
Department of Statistics
The Ohio State University
Columbus, Ohio 43210
July 1980

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
TR - 209	ADA087373	

4. TITLE (and Subtitle)	5. TYPE OF REPORT & PERIOD COVERED
HETEROSCEDASTICITY.	Technical Report,
	6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)	8. CONTRACT OR GRANT NUMBER(s)
Edward J. Dudewicz	N00014-78-C-0543

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Department of Statistics The Ohio State University Columbus, Ohio 43210	NR 042-403

11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE
Office of Naval Research Department of the Navy Arlington, Virginia 22217	July 1980
	13. NUMBER OF PAGES
	25

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report)
	Unclassified
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approval for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Heteroscedasticity, Homoscedasticity, Transformations, Unequal Variances, Inversion, Two-Stage Sampling, Heteroscedastic Method, Multiple Comparisons, ANOVA.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Recently developed methods of dealing with heteroscedasticity without transformations, in ANOVA, confidence intervals, and ranking and selection, are given. Practical aspects, a numerical example, contrasts with other approaches, and critical analysis are given.

DD FORM 1473
1 JAN 73

HETEROSCEDASTICITY

Edward J. Dudewicz, Ph. D.
Professor, Department of Statistics
The Ohio State University
Columbus, Ohio 43210, U. S. A.

If one has observed n independent and identically dis-
tributed random variables X_1, ..., X_n, it is often of in-
terest to estimate their mean $\mu = E(X_1)$ and variance
$\sigma^2 = Var(X_1)$. If X_1 has a normal distribution then

$$\bar{X} = \sum_{i=1}^{n} X_i/n \text{ and } s^2 = \sum_{i=1}^{n-1} (X_i - \bar{X})^2/(n-1)$$

. are, respectively, the usual unbiased estimators of μ and
of σ^2, and are in many senses optimal estimates. Also, it
is known that

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1},$$

i.e. $(\bar{X} - \mu)/(s/\sqrt{n})$ has Student's-t distribution with n - 1
degrees of freedom, which allows one not only to give a
point estimate of μ, but also a $100(1-\alpha)\%$ confidence inter-
val on μ:

_____

$$\bar{X} - t_{n-1,1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \le \mu \le \bar{X} + t_{n-1,1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \; .$$

A problem with this interval is that its length,

$$2 t_{n-1,1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \; ,$$

is a random variable and cannot be controlled to be (e.g.) $\le 2L$ by choice of the sample size n. The solution of this problem given below (Section 1) is the base of many solutions to problems of heteroscedasticity.

If X_1 has a non-normal distribution, then $\bar{X}$ and s^2 are still unbiased estimators of μ and σ^2, respectively. However, $\bar{X}$ and s^2 are no longer independent random variables and confidence intervals for μ (even with random length) are not generally available. Moreover, $\bar{X}$ and s^2 (while each asymptotically normal by the Central Limit Theorem) are no longer optimal estimators in this setting. Here transformations are often used, i.e. one lets $Y_i = \xi(X_i)$, $1 \le i \le n$, for some function $\xi(\cdot)$ such that $\xi(X_1)$ is normally distributed. Then $E(Y_1)$ and $Var(Y_1)$ can be estimated optimally, and confidence intervals provided. Methods are available (Section 2) for relating these estimates and intervals back to μ and σ^2, the quantities of primary interest.

If one has observations from several sources, say X_{i1}, X_{i2}, ... (which are independent and identically

distributed) from source i (i = 1, 2, ..., k), interest
is often in the means $\mu_i = E(X_{i1})$ and variances
$\sigma_i^2 = Var(X_{i1})$. Until recently, the procedures available
for these problems assumed normality and
$\sigma_1^2 = \ldots = \sigma_k^2 = \sigma^2$ and provided performance characteris-
tics (e.g. power for a test, confidence coefficient or
length for a confidence interval, probability of correct
selection for a ranking-and-selection procedure) which
depended on the unknown σ^2. The solutions given below for
these problems (Sections 3, 4, 5) do not assume equal var-
iances, and yet do allow full control of such performance
characteristics as power, confidence interval length, and
probability of correct selection.

The procedures given require the experimenter to have
design-control, but generalize to any statistical problem
via the Heteroscedastic Method (Section 6). Problems of
non-normality, comparison with the usual variance-stabili-
zing-transformation approach, and other comparisons and
questions which arise in practical implementation, are
discussed throughout.

1. Basic Sampling Rule $\mathcal{E}_B$ (n_0, w).

If we are able to observe independent and identically
distributed normal random variables $X_1, X_2, \ldots$ with mean
μ and variance σ^2 (both unknown), and wish to make

inference about μ, let us use the

Sampling Rule $\mathcal{S}_B(n_0, w)$: Take an initial sample $X_1, \ldots, X_{n_0}$ of size $n_0 \ (\geq 2)$, and calculate

$$\overline{X}(n_0) = \sum_{j=1}^{n_0} X_j / n_0, \quad s^2 = \sum_{j=1}^{n_0} (X_j - \overline{X}(n_0))^2 / (n_0 - 1),$$

$$N = \max \ \{n_0 + 1, \ [(ws)^2]\}$$

where $w > 0$ (depends on the problem under consideration) and $[y]$ denotes the smallest integer $\geq y$ (e.g. $[5.1] = 6$ ... introduced because sample sizes must be integers). Take $N - n_0$ additional observations $X_{n_0+1}, \ldots, X_N$ and calculate:

$$\overline{X}(N-n_0) = \sum_{j=n_0+1}^{N} X_j / (N-n_0),$$

$$\tilde{X} = b\overline{X}(n_0) + (1-b)\overline{X}(N-n_0)$$

where

$$b = \left(\frac{n_0}{N} \ 1 - \sqrt{1 - \frac{N}{n_0} (1 - \frac{N-n_0}{(ws)^2})} \right) ;$$

and

$$\overline{X} = \sum_{j=1}^{N} X_j / N.$$

Then Stein showed in 1945 that, for his sampling rule $\mathcal{S}$,

$$\frac{\tilde{X} - \mu}{1/w} \sim t_{n_0 - 1}.$$

Therefore,

$$\tilde{X} - t_{n_0-1,1-\frac{\alpha}{2}} \; \frac{1}{w} \leq \mu \leq \tilde{X} + t_{n_0-1,1-\frac{\alpha}{2}} \; \frac{1}{w}$$

is an exact $100(1-\alpha)\%$ confidence interval for μ, and its half-length can be fully controlled to a pre-set number $L > 0$ by taking the arbitrary $w > 0$ in $\mathcal{S}_B(n_0, w)$ such that

$$t_{n_0-1,1-\frac{\alpha}{2}} \; \frac{1}{w} = L, \; \text{i.e.} \; w = \frac{t_{n_0-1,1-\frac{\alpha}{2}}}{L}.$$

Note that, since N is an increasing function of w, the total sample size N required is larger for: smaller half-length L; as well as for higher confidence $1-\alpha$.

The above procedure is valid for any preliminary sample size $n_0 \geq 2$. Since $t_{n_0-1,1-\frac{\alpha}{2}}$ decreases as n_0 increases, it is reasonable to make n_0 large if possible. However, the decrease is negligible after $n_0 \geq 12$ or so, hence it is reasonable to take $n_0 = 12$ (or larger if one is sure to, e.g., take $n_0 \geq B$ for some positive integer $B \geq 12$).

The validity of the procedure for any n_0 (with no "optimal" choice of n_0 being obvious) bothered early workers in the field and led to dis-use of these early procedures. The above realization that $n_0 \geq 12$ is all that is required for good results in practice as far as n_0 is concerned is a factor leading to great current interest in

these procedures and their extensions (sections to follow below). One may think of the situation as follows: cne's total sample size N is approximately $w^2 s^2$, and taking n_0 very small will force a large total sample size simply be- cause of a poor initial estimate s^2. E.g., see Table 1.

2. Basic Non-Normality and Transformations.

If X_1 in Section 1 is non-normal, one often uses such transformations as

$$\xi_1(X_1) = \sqrt{X_1 - a}$$
$$\xi_2(X_1) = X_1^{1/3}$$
$$\xi_3(X_1) = \log_{10}(X_1)$$
$$\xi_4(X_1) = \arcsin \sqrt{X_1}$$
$$\xi_5(X_1) = \sinh^{-1} \sqrt{X_1}.$$

If one of these, say $\xi(X_1)$, is normally distributed, then the mean and variance of $Y_i = \xi(X_i)$ may be estimated by

$$\overline{Y} = \sum_{j=1}^{n} Y_j/n, \quad s_Y^2 = \sum_{j=1}^{n} (Y_j - \overline{Y})^2/(n-1).$$

However, interest in many cases is not in $E\xi(X_1)$ and $\mathrm{Var}\xi(X_1)$, but in the original problem units, i.e. EX_1 and $\mathrm{Var}(X_1)$. Simply using the inverse transformation, e.g. estimate EX_1 by

$$\overline{Y}^2 + a$$

in the case of ξ_1, results in a biased estimate. However, J. Neyman and E. L. Scott showed in 1960 that the unique

minimum variance unbiased estimators of $E(X_1)$ are as in Table 2. General results for second order entire functions were also given by Neyman and Scott.

In 1968 M. H. Hoyle provided the MVUE's of $Var(X_1)$ and, more importantly, of the variances of the estimators of EX_1 given in Table 2; see Table 3. These latter can be used to obtain approximate 95% confidence intervals for EX_1, e.g. when using $\sqrt{X_1-a}$:

$$\mu_{\epsilon}\bar{Y}^2+a+(1-\frac{1}{n})s_Y^2 \pm 2\sqrt{\frac{4}{n}s_Y^2\ \bar{Y}^2+s_Y^4\ \{(1-\frac{1}{n})^2-\frac{n-1}{n+1}(1-2(1-\frac{1}{n})^2+3(1-\frac{1}{n})^4)\}}.$$

Table 1. $w = t_{n_0-1, 1-\frac{\alpha}{2}} /L$ for $L = 1$, $1 - \alpha = .95$.

n_0	2	3	4	5	6	7	8	9	10	11	12	∞
w	12.706	4.303	3.182	2.776	2.571	2.447	2.365	2.306	2.262	2.228	2.201	1.960

Table 2. Transformations and MVUE's* of $E(X_1)$.

$\xi(X_1)$	MVUE of EX_1
$\sqrt{X_1 - a}$	$\bar{Y}^2 + a + (1-\frac{1}{n}) s_Y^2$
$\log_{10}(X_1)$	$10^{\bar{Y}} S((\ln 10)^2(1-\frac{1}{n})(n-1)s_Y^2 \, , \, n-1)$
$\arcsin \sqrt{X_1}$	$(\sin^2\bar{Y} - 0.5)S(4(\frac{1}{n}-1)(n-1)s_Y^2 \, , \, n-1) + 0.5$
$\sinh^{-1} \sqrt{X_1}$	$(\sinh^2\bar{Y} + 0.5) \, S(4(1-\frac{1}{n})(n-1)s_Y^2 \, , \, n-1) - 0.5$

* Here $S(a, b) = \sum\limits_{i=0}^{\infty} \frac{1}{i!} \frac{\Gamma(\frac{b}{2})}{\Gamma(i+\frac{b}{2})} \, (\frac{a}{4})^i$; this series converges faster than the series for the exponential function.

Table 3. MVUE's* of Variances of Table 2 Estimators of EX_1.

$\xi(X_1)$	MVUE of Variance of MVUE of EX_1
$\sqrt{X_1 - a}$	$\frac{4}{n}s_Y^2\,\bar{Y}^2 + s_Y^4\,\{(1-\frac{1}{n})^2 - \frac{n-1}{n+1}(1-2(1-\frac{1}{n})^2+3(1-\frac{1}{n})^4)\}$
$\log_{10}(X_1)$	$10^{2\bar{Y}}\{S^2((\ell n10)^2(1-\frac{1}{n}),n-1)-S(2(\ell n10)^2(1-\frac{2}{n}),n-1)\}$
$\arcsin\sqrt{X_1}$	$(\hat{EX}_1)^2 - \frac{1}{4}\frac{1}{8}S(-8,n-1)+\frac{1}{2}\cos(2\bar{Y})S(-4(1-\frac{1}{n}),n-1)-\frac{1}{8}\cos(4\bar{Y})S(-8(1-\frac{2}{n}),n-1)$
$\sinh^{-1}\sqrt{X_1}$	$(\hat{EX}_1)^2 - \frac{1}{4}\frac{1}{8}S(8,n-1)+\frac{1}{2}\cosh(2\bar{Y})S(4(1-\frac{1}{n}),n-1)-\frac{1}{8}\cosh(4\bar{Y})S(8(1-\frac{2}{n}),n-1)$

* For S(a, b), see footnote to Table 2.

3. Heteroscedasticity (Several Sources): Tests.

If X_{i1}, X_{i2}, ... are independent and identically dis-
tributed normal random variables with mean μ_i and variance
σ_i^2 ($i = 1, 2, \ldots, k$), then one talks of homoscedasticity
if $\sigma_1^2 = \ldots = \sigma_k^2$, and of heteroscedasticity otherwise.
Experimenters have often been cautioned that "the assump-
tion of equal variability should be investigated" (e.g. by
Cochran and Cox in 1957, by Juran, Gryna, and Bingham in
1974), but no exact statistical procedures have been avail-
able for dealing with cases where one finds that variabili-
ties are in fact unequal. (A variance-stabilizing trans-
formation is commonly employed, e.g. arcsin for binomial
data; however, if X_{ij} <u>is</u> normal, then $\xi(X_{ij})$ will be non-
normal: the transformation method has not been developed
to handle this problem except in special cases, and even
there one deals not with the parameters $\mu_1, \ldots, \mu_k$ of
basic interest if one uses such a transofrmation, but
rather with some transform whose meaning, i.e. interpreta-
bility, will not often be clear. We do not therefore re-
gard transformations as of general use in cases of $k \geq 2$
when $\mu_1, \ldots, \mu_k$ are parameters of natural interest (not
arbitrary parametrizations).)

It was first developed in the 1970's by E. J. Dudewicz
that, applying sampling procedure $\mathbf{S}_B(n_0, w)$ of Section 1
separately to each source of observations, one would obtain

the ability to fully control the performance characteristics of statistical procedures even in the presence of heteroscedasticity. Let $\tilde{X}_i$ denote the result of applying the sampling procedure to X_{i1}, X_{i2}, ... $(i = 1, 2, ..., k)$. In the case $k = 1$ one can develop procedures (as in Section 2) using $\tilde{X}$, but if one replaces this by $\bar{X}$ at the end then the procedure is still valid. In fact, it has slightly better performance characteristics (higher power), and is even simpler ($\bar{X}$ being simpler than $\tilde{X}$, the latter being a random-weighted combination of the sample means of the first and second stages of sampling). However, this improvement is not large: approximately the amount that increasing sample size from N to N + 1 will buy (one observation's worth). This improvement of $\bar{X}$ over $\tilde{X}$ has been shown to not hold in most situations where $k \geq 2$: in most such cases if $\bar{X}_1$, ..., $\bar{X}_k$ are used to replace $\tilde{X}_1$, ..., $\tilde{X}_k$ then the procedure no longer has the desired performance characteristics.

We will describe the new ANOVA procedures in the context of the one-way layout; similar procedures are available for higher-way layouts. In the one-way layout, we might want to test the null hypothesis

$$H_0: \quad \mu_1 = \mu_2 = ... = \mu_k.$$

Define

$$\tilde{F} = \sum_{i=1}^{k} w^2(\overset{\sim}{\overline{X}}_i - \overset{\sim}{\overline{X}}.)^2,$$

where

$$\overset{\sim}{\overline{X}}. = \frac{1}{k} \sum_{i=1}^{k} \overset{\sim}{\overline{X}}_i,$$

and reject H_0 if and only if

$$\tilde{F} > \tilde{F}_{k,n_0}^{\alpha},$$

where $\tilde{F}_{k,n_0}^{\alpha}$ is the upper α^{th} percent point of the null distribution of $\tilde{F}$. This null distribution is the same as that of the random variable $Q = \sum_{i=1}^{k} (t_i - \overline{t}.)^2$ where the $\{t_i\}$ are independent identically distributed Student's-t variates with n_0-1 degrees of freedom and $\overline{t}. = (1/k) \sum_{i=1}^{k} t_i$.

Values of $\tilde{F}_{k,n_0}^{\alpha}$ obtained by a Monte Carlo sampling experiment, along with the power attained at various alternatives measured by $\delta = \sum_{i=1}^{k} (\mu_i - \overline{\mu}.)^2$, for various given $1/w^2$ values, are presented in Bishop and Dudewicz (1978). There is a need for approximations to the percentage points of the $\tilde{F}$ statistics under the null and alternative distributions. Such approximations are available in the general setting (see Dudewicz and Bishop (1979)), and have been studied as to goodness in special cases (see Bishop, Dudewicz, Juritz, and Stephens (1978)). Consider first the distribution of $\tilde{F}$ as $n_0 \to \infty$. This limiting distribution is noncentral chi-square with k-1 degrees of freedom

and noncentrality parameter $\Delta = \sum\limits_{i=1}^{k} w^2(\mu_i - \bar{\mu}.)^2$, denoted by $\chi^2_{k-1}(\Delta)$. However numerical results indicate that for small n_0 the tails of this limiting distribution are too light to give a good approximation. One therefore approximates by a $((n_0-1)/(n_0-3)) \cdot \underset{\sim}{\chi^2_{k-1}}(\Delta)$ random variable (in which case F and its approximating distribution have the same expected value under H_0).

Let us illustrate with a numerical example. Suppose we wish to test the hypothesis that 4 different chemicals are equivalent in their effects. Suppose we decide to take initial samples of size 10 with each treatment, want only a 5% chance of rejecting H_0 if in fact H_0 is true, and want an 85% chance of rejecting H_0 if the spread among μ_1, μ_2, μ_3, μ_4 is at least 4.0 units. We then proceed, step by step, as follows.

Step 1. (Problem specification.) Here there are $k = 4$ sources of observations, we desire an $\alpha = .05$ level test of H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$, and if the spread among μ_1, μ_2, μ_3, μ_4 is $\delta = 4.0$ units or more we desire power (probability of then rejecting the false hypothesis H_0) of at least $P* = .85$.

Step 2. (Choice of procedure.) Assuming we do not know that $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$, only procedure $\mathfrak{B}_B(n_0, w)$ can guarantee the specifications. It requires we sample n_0 observations in our first stage, and recommends n_0

be at least 12 (though any $n_0 \geq 2$ will work). Suppose the experimenter only wants to invest 40 units in first-stage experimentation and sets $n_0 = 10$.

Step 3. (First stage.) Draw $n_0 = 10$ independent observations from each source, with results as in Table 4.

Table 4. First Stage Samples.

Chemical 1	Chemical 2	Chemical 3	Chemical 4
77.199	80.522	79.417	78.001
74.466	79.306	78.017	78.358
82.746	81.914	81.596	77.544
76.208	80.346	80.802	77.364
82.876	78.385	80.626	77.554
76.224	81.838	79.011	75.911
78.061	82.785	80.549	78.043
76.391	80.900	78.479	78.947
76.155	79.185	81.798	77.146
78.045	80.620	80.923	77.386

Step 4. (Analysis of first stage data.) We now calculate the first stage sample variances s_1^2, s_2^2, s_3^2, s_4^2, the total sample sizes needed from the four sources N_1, N_2, N_3, N_4, and the factors b_1, b_2, b_3, b_4 to be used in the second stage analysis. The s_i^2's are given in Table 5, along with the other quantities. The w needed is found as follows.

We desire power $P* = .85$ (Step 1 above) when

$$\Delta = \frac{w^2\delta^2}{4} = \frac{w^2(4.0)^2}{4} = 4.0w^2 .$$

To set w for this power requirement, we first need to know "When do we reject?". We know we will later reject H_0 if $\tilde{F} > \tilde{F}_{4,10}^{.05}$ where, approximately,

$$\tilde{F}_{4,10}^{.05} = \frac{n_0-1}{n_0-3} (7.81) = 10.04.$$

The 7.81 is the value a central chi-square random variable with $k - 1 = 4 - 1 = 3$ degrees of freedom exceeds with probability $\alpha = .05$ (see standard tables, e.g., p. 137 of Pearson and Hartley (1970) or p. 459 of Dudewicz (1976)).

The power will be, approximately,

$$P[\chi_3^2 (\Delta) > 7.81] = .85$$

if (see p. 53 of the tables in Haynam, Govindarajulu, and Leone (1970))

$$\Delta = 12.301 ,$$

so

$$w^2 = \frac{12.301}{4.0} = 3.075.$$

Table 5.   Analysis of First Stage.

	Chemical 1	Chemical 2	Chemical 3	Chemical 4
n_0	10	10	10	10
Sample Mean	77.837	80.580	80.122	77.625
s_i^2	7.9605	1.8811	1.7174	.6762
w	1.754	1.754	1.754	1.754
N_i	25	11	11	11
b_i	.330	.936	.939	.969

Step 5. (Second stage.) Draw $N_i - n_0$ observations from source i (i = 1, 2, 3, 4), yielding Table 6.

Table 6.   Second Stage Samples.

Chemical 1	Chemical 2	Chemical 3	Chemical 4
82.549	79.990	80.315	78.037
78.970			
78.496			
78.494			
80.971			
80.313			
76.556			
80.115			
78.659			
77.697			
80.590			
79.647			
82.733			
80.552			
79.098			

Step 6. (Final calculations.) We now calculate the $\tilde{X}_i$ and $\tilde{F}$, and find

$$\tilde{X}_1 = 79.079, \; \tilde{X}_2 = 80.688, \; \tilde{X}_3 = 80.197, \; \tilde{X}_4 = 77.597$$

$$\tilde{X}. = 79.390 ,$$

$$\tilde{F} = 17.38 .$$

Step 7. (Final Decision.) Since $\tilde{F} = 17.33$ exceeds $\tilde{F}_{4,10}^{.05} = 10.04$, we reject the null hypothesis and decide the chemicals differ.

4. Heteroscedasticity (Several Sources): Confidence Intervals.

The case of a confidence interval with $k = 1$ mean was considered in Section 1. When $k = 2$, a two-sided confidence interval of half-length $L > 0$ and with confidence coefficient $1-\alpha$ is given by

$$(\tilde{X}_1 - \tilde{X}_2) - L \leq \mu_1 - \mu_2 \leq (\tilde{X}_1 - \tilde{X}_2) + L$$

if we choose (in $\mathbf{S}_B (n_0, w)$)

$$w = \frac{c_{1-\frac{\alpha}{2}} (n_0)}{L}$$

where c is tabled in Table 7.

Table 7. $c_{1-\gamma}(n_0)$.

$1-\gamma$ \ n_0	10	11	12	13	14	15	20	25	30
.75	1.03	1.02	1.02	1.01	1.01	1.00	.99	.98	.98
.80	1.29	1.28	1.27	1.26	1.26	1.25	1.24	1.23	1.22
.85	1.60	1.59	1.57	1.56	1.56	1.55	1.53	1.51	1.51
.90	2.00	1.98	1.96	1.95	1.94	1.93	1.90	1.88	1.87
.95	2.61	2.58	2.56	2.53	2.52	2.50	2.45	2.42	2.41
.975	3.18	3.13	3.09	3.06	3.04	3.02	2.95	2.91	2.88
.99	3.89	3.82	3.76	3.71	3.67	3.64	3.54	3.48	3.45
.995	4.41	4.31	4.24	4.18	4.13	4.09	3.96	3.89	3.85
.999	5.61	5.45	5.32	5.22	5.14	5.07	4.86	4.74	4.67

Note that the corresponding test solves the Behrens-Fisher problem exactly in two stages, with controlled level and power.

For $k > 2$, multiple-comparison procedures are also available for many of the usual multiple-comparison confidence interval goals.

5.   Heteroscedasticity (Several Sources):   Ranking and Selection

Here $k \geq 2$ and we wish to select (indifference-zone formulation) that source with mean value max $(\mu_1, \ldots, \mu_k)$. Let $\mu_{[1]} \leq \cdots \leq \mu_{[k]}$ denote the ordered values of $\mu_1, \ldots, \mu_k$; thus $\mu_{[k]}$ denotes max $(\mu_1, \ldots, \mu_k)$, etc..

The performance characteristic of interest is the probability that we will make a correct selection (CS), i.e. that the population actually selected is the one which has mean $\mu_{[k]}$. Following Bechhofer (1954), we require that our P(CS) have at least a specified value P^* ($1/k < P^* < 1$) whenever the largest mean is at least δ^* ($0 < \delta^*$) more than the next-to-largest mean; i.e. we require

$$P(CS) \geq P^* \text{ whenever } \mu_{[k]} - \mu_{[k-1]} \geq \delta^*.$$

This problem was considered by Dudewicz and Dalal (1975). The procedure is to select that source which yields the largest of $\tilde{X}_1, \ldots, \tilde{X}_k$; i.e.

$$\text{Select } \pi_i \text{ iff } \tilde{X}_i = \max(\tilde{X}_1, \ldots, \tilde{X}_k).$$

In the sampling rule $\mathfrak{s}_B(n_0, w)$ one chooses

$$w = \frac{c_{P^*}(n_0)}{\delta^*}$$

where $c_{P^*}(n_0)$ (for specified values of P^* and n_0) is tabled in Table 7 for $k = 2$, and is tabled in Dudewicz, Ramberg, and Chen (1975) for $k > 2$. Approximations for $k > 25$ are given by Dudewicz and Dalal (1975). Subset-selection procedures are also available there.

6.   The Heteroscedastic Method.

The above special-case solutions have been placed into a general theory with the Heteroscedastic Method of Dudewicz and Bishop (1979). In a general decision-theoretic setting, they show how to develop procedures like the above in any problem. It is also shown that no single-stage procedure can solve most such problems.

Some questions one might ask about the procedures thus produced are as follows. First, how do they perform under violation of normality? Here, Iglehart (1977) has shown, in some computational settings, that replacing s^2 by a jack-knife estimator is sufficient to preserve the main proper-ties of the procedures. Other recent work (Dudewicz and van der Meulen (1980)) shows asymptotic validity under asymptotic normality. Second, are they preferable to com-parable sequential procedures? Here it should be noted that in most cases there are no "comparable" sequential pro-cedures: the sequential procedures of Chow-Robbins type usually mentioned only have asymptotic validity even under exact normality, while the $\mathbf{3}_B(n_0, w)$-based two-stage proce-dures have exact known properties. It is sometimes claimed that the sequential procedures are more efficient, but this is only as (e.g.) $\sigma_i^2 \to 0$. The so-called inefficiency of $\mathbf{3}_B(n_0, w)$ in this situation is due to the fact it then requires $N = n_0+1$ (since $N \geq n_0+1$ always) and in fact (as

$\sigma_i^2 \rightarrow 0$) an $N \rightarrow 1$ will suffice. This appears to have little practical relevance, as one usually knows trivial sample sizes will be insufficient for one's problems; it is rather a curiosity of mathematical interest only.

As a final note, we mention that while variance-stabilizing transformations and other approximate methods have existed for many years, most experimental situations are such that the problem is far from solved by these approximate methods. For example, such methods misallocate sample size by taking the same sample size from a treatment with relatively small variability, as from a treatment with relatively large variability, even though the need for observations on the latter is substantially greater and they have a greater beneficial effect on performance characteristics of the overall analysis. Also, procedures based on $S_B(n_0, w)$ behave acceptably even if variances are equal, hence the equality of variances tests, which are known to be weak in power, can be skipped and these new procedures applied directly without regard to equality or inequality of variances.

Annotated References.

Bechhofer, R. E. (1954): "A single-sample multiple deci-
sion procedure for ranking means of normal populations
with known variances," Annals of Mathematical
Statistics, Vol. 25, pp. 16-39. [The original paper
on ranking and selection methods.]

Bishop, T. A. and Dudewicz, E. J. (1978): "Exact analysis
of variance with unequal variances: test procedures
and tables," Technometrics, Vol. 20, pp. 419-430.
[The original ANOVA procedures for heteroscedastic
situations, with tables and approximations needed for
implementation.]

Bishop, T. A., Dudewicz, E. J., Juritz, J. and Stevens,
M. A. (1978): "Percentage points of a quadratic form
in Student-t variates," Biometrika, Vol. 65, pp. 435-
439. [Considers approximating the F distribution.]

Chapman, D. G. (1950): "Some two sample tests," Annals of
Mathematical Statistics, Vol. 21, pp. 601-606. [Con-
sidered the k = 2 test of H_0: $\mu_1 = \mu_2$ vs.
H_1: $|\mu_1 - \mu_2| = d$, and also H_0: $\mu_1 = r \mu_2$. Tabled
c for $n_0 = 2(2)12$ and $1-\alpha = .975, .995$, correct to
0.1 (except for a gross error when $n_0 = 4$).]

Cochran, W. G. and Cox, G. M. (1957): Experimental Designs
(Second Edition), John Wiley & Sons, Inc., New York.
[On p. 91, note the need to test for heteroscedasticity.]

Dantzig, G. B. (1940): "On the non-existence of tests of
"Student's" hypothesis having power functions independent
of σ," Annals of Mathematical Statistics, Vol. 11,
pp. 186-191. [First to show that one stage procedures
could not solve many practical problems.]

Dudewicz, E. J. (1976): Introduction to Statistics and
Probability, American Sciences Press, Columbus, Ohio.

Dudewicz, E. J. and Bishop, T. A. (1979): "The heteros-
cedastic method," Optimizing Methods in Statistics,
edited by J. S. Rustagi, Academic Press, New York,
pp. 183-203. [Develops the Heteroscedastic Method as
a unifying procedure in a general setting, and shows
how the procedures referred to here fit in as special
cases.]

Dudewicz, E. J. and Dalal, S. R. (1975): "Allocation of
    observations in ranking and selection with unequal
    variances," Sankhyā, Vol. 37B, pp. 28-78. [Solves
    the heteroscedastic ranking and selection problem
    in indifference-zone and subset-selection settings.
    Gives extensive tables and suggestions on solutions
    of other problems with similar methods.]

Dudewicz, E. J., Ramberg, J. S. and Chen, H. J., (1975):
    "New tables for multiple comparisons with a control
    (unknown variances)," Biometrische Zeitschrift, Vol.
    17, pp. 13-26. [Gives procedures and theory for one-
    sided multiple-comparisons with a control, plus
    extensive tables of $c_{1-\gamma}(n_0)$ useful in many problems.]

Dudewicz, E. J. and van der Meulen, E. C. (1980): "Entropy-
    based statistical inference, II: Selection-of-the-
    best/complete ranking for continuous distributions on
    (0,1), with applications to random number generators,"
    Communication No. 123, Mathematical Institute,
    Katholieke Universiteit Leuven, Belgium. [New results
    on validity of $S_B(n_0, w)$ under asymptotic (rather than
    exact) normality.]

Haynam, G. E., Govindarajulu, Z. and Leone, F. C. (1970):
    "Tables of the cumulative non-central chi-square dis-
    tribution," Selected Tables in Mathematical Statistics,
    Vol. I, edited by H. L. Harter and D. B. Owen,
    Markham Publishing Company, Chicago, pp. 1-78.

Iglehart, D. L. (1977): "Simulating stable stochastic
    systems, VII: Selecting the best system," TIMS
    Studies in the Management Sciences, Vol. 7, pp. 37-49.
    [Suggested using a jackknife variance estimator with
    $S_B(n_0, w)$, and indicated it solves non-normality
    problems in his context.]

Johnson, N. L. and Kotz, S. (1970): Distributions in
    Statistics: Continuous Univariate Distributions - 2,
    Houghton Mifflin Company, Boston.

Juran, J. M., Gryna, F. M., Jr., and Bingham, R. S., Jr.
    (Editors) (1974): Quality Control Handbook (Third
    Edition), McGraw-Hill, Inc., New York. [Recommend
    testing for heteroscedasticity on p. 46 of Section 27.]

Miller, R. G. (1974): "The jackknife - a review," _Biometrika_, Vol. 61, pp. 1-15. [Recent survey of jackknife methods, with an extensive bibliography.]

Pearson, E. S. and Hartley, H. O. (Editors) (1970): _Biometrika Tables for Statisticians_, Vol. I (Third Edition, Reprinted with additions), Cambridge University Press, London.

Ruben H. (1962): "Studentisation of two-stage sample means from normal populations with unknown variances. I. General theory and application to the confidence estimation and testing of the difference in populations means," _Sankhyā_, Series A, Vol. 24, pp. 157-180. [Looks at testing H_0: $\mu_1 = \mu_2$ when k = concentrating attention on $\overline{X}_1$ and $\overline{X}_2$... hence missing the generalizations found in the 1970's.]

Stein, C. (1945): "A two-sample test for a linear hypothesis whose power is independent of the variance," _Annals of Mathematical Statistics_, Vol. 16, pp. 243-258. [The original reference to $S_B(n_0, w)$, but did not consider heteroscedasticity, perhaps due to emphasis on $\tilde{X}$ as a replacement for $\overline{X}$.]